Chapter 2
Technical Aspects of Geostatistics

2-1. General

a. This chapter provides technical aspects or the necessary theoretical background for understanding kriging applications. Emphasis will be placed on presentation of the basic ideas; long formulas or derivations are kept to a minimum. Statistical terms that are commonly used in geostatistical applications will be highlighted with bold text and briefly defined as they are introduced; notation used in this ETL is also tabulated in Appendix B. The reader who wishes a more thorough discussion of these fundamental concepts may consult the references cited in Chapter 3. Previous exposure to engineering statistics at the level of Devore (1987) and Ross (1987) would be helpful in understanding some parts of this chapter. Readers with limited statistical experience may wish to briefly scan this chapter and refer back to it after reading the remaining chapters.

b. In section 2-2, regionalized random variables are discussed. Regionalized random variables constitute the random process that is sampled to obtain the observed data that are available for analysis. Basic ideas related to probability distributions, means, variances, and correlation are introduced. The variogram, which is the fundamental tool used in geostatistics to analyze spatial correlation, is introduced in section 2-3. In section 2-4 how kriging is used to obtain the best weights for spatial prediction is discussed, and how the mean squared prediction error for these predictions is computed is also shown. Section 2-5 deals briefly with co-kriging, which is prediction of one variable based not only on measurements of that variable but on measurements of other variables as well. Finally, section 2-6 shows how kriging may be applied to determine not just optimal spatial predictions but also probabilities associated with various events, such as extreme events that may be of importance in risk-based analyses.

2-2. Regionalized Random Variables

a. General.

(1) Suppose the extent of groundwater contamination of a particular pollutant over a given study area is being determined. To simplify the presentation, all data are assumed to be distributed over a two-dimensional region. In three-dimensional groundwater flow systems, one could study the depth-averaged concentration of a pollutant or the concentration of the pollutant in a particular horizontal stratum of the flow system. Let a vector $\underline{x}=(u,v)$ denote an arbitrary spatial location in the study area. Unless otherwise stated, it will be assumed throughout the ETL that u is the east-west coordinate and v is the north-south coordinate (Figure 2-1). Denote by $z(\underline{x})$ a measurement at location $\underline{x}$, such as the concentration of a pollutant. The ultimate goal of an investigator would be to determine $z(\underline{x})$ for all locations in the study area. However, without explicit knowledge of the flow and transport field, this goal cannot be achieved. Therefore, suppose, instead, that the goal is to estimate the values of $z(\underline{x})$ with a given error tolerance. In other situations, small estimation error over some parts of the study area (for instance, near a domestic water supply) may need to be obtained, while allowing larger estimation errors in other parts of the study area. The theory of regionalized random variables is designed to accomplish these goals.

(2) In the regionalized random variable theory, the true measurement $z(\underline{x})$ is assumed to be the value of a **random variable** $Z(\underline{x})$. Associating a random variable $Z(\underline{x})$ with a true measurement $z(\underline{x})$ is done for the purpose of characterizing the degree of uncertainty in the quantity of interest at point $\underline{x}$. If there is no actual measurement taken at $\underline{x}$, then the values taken on by $Z(\underline{x})$ represent "potential" measurements at $\underline{x}$; that is, $Z(\underline{x})$ represents possible values that might be expected if a measurement were taken at $\underline{x}$. Because there is uncertainty associated with $Z(\underline{x})$, it needs to be characterized by a **probability distribution**, defined by $P[Z(\underline{x}) \leq c]$ where P denotes probability and c is any constant.

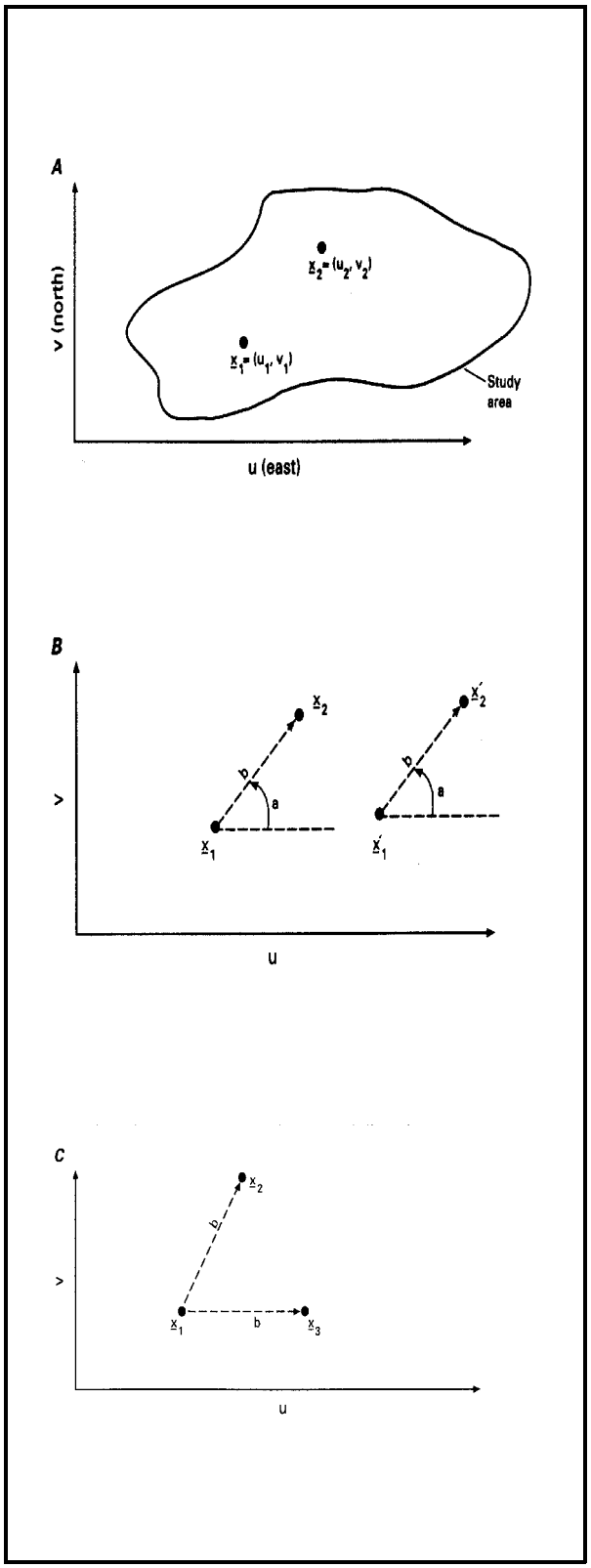


Figure 2-1. Diagrams showing A, hypothetical study area; B, stationary covariance function; and C, isotrophic covariance function

This distribution is a function of *c*, and, to be completely defined, needs to be known for all values of *c*. The distribution is used to make evaluations such as: suppose that we have no measurement of concentration of a certain contaminant at $\underline{x}$, but the distribution is known, and a threshold value of c = 8 mg/l is of interest. If $P\ [Z\ (\underline{x}) \leq 8]\ =\ 0.60$, then, if a measurement were made at $\underline{x}$, there is a 60-percent chance of obtaining a value less than or equal to 8 mg/l. The distribution also may be used to calculate other probabilities, such as the probability of obtaining a value in some specified interval.

(3) An important concept to keep in mind in all geostatistical applications is the **support** of the regionalized random variable. The support of $Z(\underline{x})$ is the in situ geometric unit represented by an individual sample. For example, in a soil contamination study, sample $Z(\underline{x})$ might represent the concentration of a contaminant in a vertical soil core 0.1 m in diameter and 1 m in length, and centered at location $\underline{x}$. Thus, even though $Z(\underline{x})$ is defined at a particular point, it is representative of a volume of soil. Changing the support of $Z(\underline{x})$ will usually change its probability distribution. Therefore, the observations in a geostatistical analysis should all have the same support. The method called point, or punctual, kriging, described in section 2-4, is designed to predict values of $Z(\underline{x})$ with the same support as the sample data.

(4) A concept closely related to support is that of **estimation block**, which is a geometric unit larger than the support of a single observation, for which a single representative value is desired. For example, in the above soil contamination study, it may be necessary to estimate the average concentration of the contaminant in a truckload of soil excavated from a block 6 m long, 6 m wide, and 0.3 m thick. Using a method called block kriging, also described in section 2-4, the block average can be predicted based on individual measurements.

(5) Although the distribution of $Z(\underline{x})$ completely characterizes $Z(\underline{x})$ at any particular location, this distribution indicates nothing about the relations among the values of $Z(\underline{x})$ at different

locations, which is very important, because geostatistics is based on using a measurement of a regionalized variable at one location to gain information about values of the variable at another location. The notion of distribution of $Z(\underline{x})$ at a single location is readily generalized to two or more locations. For two locations, if we let $\underline{x}_1$ and $\underline{x}_2$ be two distinct locations, then the **joint probability distribution** is defined to be the probability $P[Z(\underline{x}_1) \le c_1, Z(\underline{x}_2) \le c_2]$ for any constants c_1 and c_2. This latter probability means the probability that both $Z(\underline{x}_1) \le c_1$ and $Z(\underline{x}_2) \le c_2$. If the variables $Z(\underline{x}_1)$ and $Z(\underline{x}_2)$ are statistically independent of one another, then the joint probability distribution can be obtained as the product of the individual probability distributions,

$$P[Z(\underline{x}_1) \le c_1, Z(\underline{x}_2) \le c_2] = P[Z(\underline{x}_1) \le c_1] P[Z(\underline{x}_2) \le c_2] \quad (2\text{-}1)$$

However, in most applications, $Z(\underline{x}_1)$ and $Z(\underline{x}_2)$ will not be statistically independent and their joint distribution cannot be obtained from the individual distributions. When this joint distribution description is applied to more than two locations, specification of the full spatial distribution of Z would require knowing the joint distribution of $Z(\underline{x}_1)$, ..., $Z(\underline{x}_n)$ for any set of n spatial locations and for any n; however, except in very special cases, working with the full set of distribution functions of $Z(\underline{x})$ is not feasible and is not done.

(6) To simplify the problem even further, various parameters of the distributions are usually considered rather than dealing with the entire distributions. The parameter most commonly used to characterize a distribution is the **mean**, or, because the mean in geostatistical applications depends on the spatial variable $\underline{x}$, the mean may be called the **spatial mean,** or the drift. In statistics, the mean is referred to as the expectation (E) of the random variable $Z(\underline{x})$, and the symbol m is used in this report to denote this expectation Thus,

$$\mu(\underline{x}) = E[Z(\underline{x})] \quad (2\text{-}2)$$

is used to denote the mean, or expected value, of the bracketed term, in this case $Z(\underline{x})$. It is intuitively helpful to think of the expectation as an average. In fact, if the distribution of $Z(\underline{x})$ assigned equal probability to a finite number of values, then the expectation of $Z(\underline{x})$ would indeed be the simple average of these numbers. In geostatistics, however, $Z(\underline{x})$ is usually assumed to take on any value in a continuous range of possible values, rather than being limited to a discrete set of values. In this case, calculus needs to be used to define the expectation. The following example illustrates the difference between averages and expectations.

b. Example 1.

(1) An experiment consists of injecting a conservative tracer at a particular well in a steady-state groundwater flow system and measuring the concentration, $Z_1(\underline{x})$, of the tracer in a neighboring well 24 hr later. The tracer is then allowed to flush from the system, and the experiment is repeated a second time to obtain another concentration measurement, $Z_2(\underline{x})$, at the same location. If this process is repeated n times, n concentration measurements $Z_1(\underline{x})$, $Z_2(\underline{x})$, ..., $Z_n(\underline{x})$ would be obtained, all at location $\underline{x}$. The average concentration level at location $\underline{x}$ is

$$\overline{Z}_n(x) = \frac{1}{n}\left(Z_1(\underline{x}) + Z_2(\underline{x}) + \ldots + Z_n(\underline{x})\right) \quad (2\text{-}3)$$

which would change depending on n and on the actual values obtained for $Z_1(\underline{x})$, $Z_2(\underline{x})$, ..., $Z_n(\underline{x})$. However, in the limit as n increases, $\overline{Z}_n(\underline{x})$ becomes closer and closer to the true mean, or expected, concentration $\mu(\underline{x})$:

$$\overline{Z}_n(\underline{x}) \rightarrow \mu(\underline{x}) \text{ as } n \text{ increases} \quad (2\text{-}4)$$

This theoretical limit is a constant value, or **population parameter**, as opposed to $\overline{Z}_n(\underline{x})$, which is a random variable, or a property of the particular sample that is taken.

(2) In example 1, no assumptions were needed concerning whether the mean changed with spatial location, because all sampling was done at one sampling location $\underline{x}$. In most HTRW applications, the mean will probably change depending on the sampling location. In addition, usually only one observation is available at any particular location. Therefore some assumptions regarding the structure of $\mu(\underline{x})$ must be made. For example, it is sometimes appropriate to assume $\mu(\underline{x}) = \mu$ is constant for all $\underline{x}$, in which case $Z(\underline{x})$ is said to have a **stationary mean**. Data which have no underlying trend such as hydraulic conductivity in a homogeneous aquifer, for example, might be assumed to have a constant mean. If the mean is constant, it makes sense to estimate it with the sample average of n observations taken at different spatial locations $\underline{x}_1, \underline{x}_2, ..., x_n$

$$\overline{Z}_n = \frac{1}{n}\left[Z(\underline{x}_1) + Z(\underline{x}_2) + ... + Z(\underline{x}_n)\right] \quad (2\text{-}5)$$

However, in contrast to example 1, $\overline{Z}_n$ defined in this way may not get closer to μ as n gets large. Because of the possible spatial correlation in the data, the size of the sampling region must be large in relation to the correlation length in order for $\overline{Z}_n$ to accurately estimate μ.

(3) In addition to the mean of $Z(\underline{x})$, its variability or dispersion is also of interest, and this variability is most commonly measured by the **(spatial) variance**, defined to be the mean of squared deviations of $Z(\underline{x})$ from $\mu(\underline{x})$ and denoted by $\sigma^2(\underline{x})$.

$$\sigma^2(\underline{x}) = E\left[(Z(\underline{x}) - \mu(\underline{x}))^2\right] \quad (2\text{-}6)$$

The **(spatial) standard deviation** $\sigma(\underline{x})$ is the square root of the variance. The following example illustrates the difference between the population variance, which has been defined above, and a sample variance.

c. Example 2.

(1) If the scenario presented in example 1 is again used, the sample variance $S_n^2(\underline{x})$ of the n measurements could be computed as follows:

$$S_n^2(\underline{x}) = \frac{1}{n-1}\sum_{i=1}^{n}\left(Z_i(\underline{x}) - \overline{Z}_n(\underline{x})\right)^2 \quad (2\text{-}7)$$

This number gives a measure of dispersion of the $Z_i(\underline{x})$ values from their sample mean . The sample variance depends on n and on the particular values observed for $Z_1(\underline{x}), Z_2(\underline{x}), ..., Z_n(\underline{x})$. However, in the limit as n increases, $S_n^2(\underline{x})$ gets closer and closer to a constant value, which is denoted by $\sigma^2(\underline{x})$. In this case, $\sigma^2(\underline{x})$ is a population parameter, and $S_n^2(\underline{x})$ is a random variable.

(2) The mean and variance can both be calculated from the probability distribution of $Z(\underline{x})$. Again, in geostatistics, the relations among regionalized variables at different locations are of interest. From the joint distribution of $Z(\underline{x}_1)$ and $Z(\underline{x}_2)$ the **(spatial) covariance function**,

$$C(\underline{x}_1, \underline{x}_2) = E\left[\left(Z(\underline{x}_1) - \mu(\underline{x}_1)\right)\left(Z(\underline{x}_2) - \mu(\underline{x}_2)\right)\right] \quad (2\text{-}8)$$

may be obtained. This function has a key role in geostatistical analyses. It is a measure of association between values obtained at point $\underline{x}_1$ and those obtained at point $\underline{x}_2$. If values at these two spatial locations tend to be greater than average or less than average at the same time, then the covariance will be positive. However, if the values vary in the opposite direction (that is, one tends to be larger than average when the other is less than average, and vice versa), the covariance will be negative.

(3) Because $C(\underline{x}_1,\underline{x}_2)$ is an unknown population parameter, it too must be estimated using a statistic computed from sample data. To make this possible, it is often assumed that the covariance function depends only on the distance between points, which is defined as the **lag** h, and not on their relative location or orientation,

$$C(\underline{x}_1, \underline{x}_2) = C(h), \qquad h = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2} \tag{2-9}$$

Under this assumption, *C(h)* can be estimated by pooling all pairs of observations that are approximately *h* units apart and computing a **sample covariance function**

$$\hat{C}(h) = \text{average}\left\{\left(Z(\underline{x}_i) - \overline{Z}\right)\left(Z(\underline{x}_j) - \overline{Z}\right) : h - \Delta h < h_{ij} < h + \Delta h\right\} \tag{2-10}$$

where h_{ij} is the distance between $\underline{x}_i$ and $\underline{x}_j$ and the average is over all pairs of points such that h_{ij} is between h-Δh and h+Δh. The distance h is called the lag and Δh is called the **lag tolerance**. There are more effective ways to estimate *C(h)* other than using Equation 2-10; for example, see Isaaks and Srivastava (1989). However, because the emphasis in this ETL is on the variogram (to be defined below) rather than the covariance function, we will not need to use the estimated covariance function.

(4) A covariance function is called **stationary** if it does not depend on the origin of the coordinate system, that is,

$$C(\underline{x}_1 + \underline{b}, \underline{x}_2 + \underline{b}) = C(\underline{x}_1, \underline{x}_2) \tag{2-11}$$

for any given vector, $\underline{b}$ (Figure 2-1). The covariance function (Equation 2-9) is stationary because changing the origin does not change the distance between the points. Substituting $\underline{x}_1 = \underline{x}_2 = \underline{x}$ in Equation 2-9 yields

$$C(\underline{x}, \underline{x}) = C(0) \tag{2-12}$$

which, combined with the definitions in Equations 2-6 and 2-8, becomes

$$\sigma^2(\underline{x}) = C(0) \text{ for all } \underline{x} \tag{2-13}$$

Therefore, when $Z(\underline{x})$ has a stationary covariance function, the variance of $Z(\underline{x})$ is constant for all $\underline{x}$. The covariance function can then be standardized by dividing it by the variance. The resulting dimensionless function of *h* is called the **spatial correlation function**,

$$\rho(h) = \frac{C(h)}{C(0)} \tag{2-14}$$

The correlation function is a scale-independent measure of linear association between values of *Z* at different locations. The spatial correlation is always between -1 and +1, with a value of zero indicating no linear association.

(5) In addition to being stationary, the covariance function in Equation 2-9 has another important property. It is also **isotropic**, or **omni-directional**, because it does not depend on the direction between the two locations. In many HTRW applications, the correlation between values of *Z* at two locations is a function of direction as well as lag. For example, contaminant concentrations in a groundwater flow system might be more highly correlated along a transect in the direction of flow than along a transect perpendicular to the flow. In that case, the covariance function depends on both the lag *h* and the angle *a* between locations,

$$C(\underline{x}_1, \underline{x}_2) = C(h, a), \qquad h = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2}, \qquad a = \text{atan}\left(\frac{v_2 - v_1}{u_2 - u_1}\right) \tag{2-15}$$

Here, *a* is the angle measured counterclockwise from the east direction (Figure 2-1). In many geostatistical publications or computer packages, the angle may be defined as clockwise from the north direction, so care should be taken in defining the appropriate angle in any application. A covariance function satisfying Equation 2-15 is called **anisotropic**, or **multi-directional.**

(6) To summarize, the basic model framework that will be used throughout the ETL is the following: the value of a measurement $z(\underline{x})$ (concentration, porosity, hydraulic head, and so on) at location $\underline{x}$ of a two-dimensional region is the value of a regionalized random variable, $Z(\underline{x})$, with mean $\mu(\underline{x})$ and stationary covariance function $C(h,a)$. Other assumptions may be added in the applications sections to analyze specific data sets, but this framework will be the basic framework from which many of the results will be derived. In some situations, the covariance stationarity assumption may be relaxed, for instance, when using the linear variogram described in the next section.

2-3. Variograms

a. Regionalized random variables differ from classical (ordinary least-squares) regression models in that the **residuals**, defined as the deviations of the regionalized random variable from its mean and denoted by

$$Z^*(\underline{x}) = Z(\underline{x}) - \mu(\underline{x}) \qquad (2\text{-}16)$$

are related to one another, whereas the residuals in a regression model are generally assumed to be independent. Thus, in the regionalized random-variable model, observed values of the residuals from sampled locations contain valuable information when predicting the value of $Z(\underline{x})$ at unsampled sites. The relationship among the residuals can be understood by examining the variogram, which is a tool that is widely used in geostatistics for modeling the degree of spatial dependence in a regionalized random variable. Although the variogram is closely related to the covariance function, there are some important differences between the variogram and covariance function that will be described below. The covariance function, and related correlation function, are more commonly used in basic statistics courses than the variogram, so many readers may be more familiar with the former concepts. However, the variogram is more widely used in geostatistics, and because of this we will adopt the variogram as the primary tool for analyzing spatial dependence in the remainder of this ETL.

b. As was the case with the covariance function, it is necessary to distinguish between the theoretical variogram, which is a population parameter, and the sample variogram, which is an estimator of the theoretical variogram obtained from observed data. The **theoretical variogram** of a regionalized random variable, $\gamma(\underline{x}_1, \underline{x}_2)$ is defined as one half of the variance of the difference between residuals at locations $\underline{x}_1$ and $\underline{x}_2$:

$$\gamma(\underline{x}_1, \underline{x}_2) = \frac{1}{2}\mathrm{Var}\,[Z^*(\underline{x}_1) - Z^*(\underline{x}_2)] \qquad (2\text{-}17)$$

Because the residuals have been mean-centered, as shown in Equation 2-16, they have a mean of zero. Therefore, using the well-known formula for the variance of a random variable X

$$\mathrm{Var}\,(X) = \mathrm{E}\,(X^2) - (\mathrm{E}X)^2 \qquad (2\text{-}18)$$

it is seen that Equation 2-17 is equivalent to

$$\gamma(\underline{x}_1, \underline{x}_2) = \frac{1}{2}\,\mathrm{E}\,[Z^*(\underline{x}_1) - Z^*(\underline{x}_2)]^2 \qquad (2\text{-}19)$$

The theoretical variogram is always non-negative, with a small value of g indicating that the residuals at locations $\underline{x}_1$ and $\underline{x}_2$ tend to be close and a large value of λ indicating that the residuals tend to be different. Equation 2-19 is sometimes called a **semi-variogram**, because of the multiplication by ½, but will be referred to in this ETL as a variogram.

c. It would be ideal to know the theoretical variogram before taking observations, but unfortunately, it must be estimated using sample data. To facilitate variogram estimation, it is usually assumed in a similar manner to the covariance function that γ depends only on the lag,

$$\gamma\ (\underline{x}_1,\ \underline{x}_2) = \gamma\ (h),$$
$$h = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2} \qquad (2\text{-}20)$$

or possibly, on the lag and angle between locations

$$\gamma\ (\underline{x}_1,\ \underline{x}_2) = \gamma\ (h,\ a),$$
$$h = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2}, \qquad (2\text{-}21)$$
$$a = a\tan\left(\frac{v_2 - v_1}{u_2 - u_1}\right)$$

(Figure 2-1). Equation 2-20 is called an **isotropic variogram** and Equation 2-21 is a **directional variogram** at angle *a*.

d. For the isotropic case, the **sample**, or **empirical, variogram** is obtained by averaging the square of all computed differences between residuals separated by a given lag:

$$\hat{\gamma}\ (h) = \frac{1}{2}\ \text{ave}\ \left\{ \left(Z^*\ (\underline{x}_1) - Z^*\ (\underline{x}_2)\right)^2 :\ h - \Delta h < h_{ij} < h + \Delta h \right\} \qquad (2\text{-}22)$$

where, as before, h_{ij} is the distance between $\underline{x}_i$ and x_j. For a given h as more and more points separated by distance $h \pm \Delta h$ are sampled and as Δh gets small, $\hat{\gamma}\ (h)$ should approach the theoretical variogram. More detail on variogram estimation will be presented in Chapter 4, including the directional case. In this section, it will be sufficient to describe some general properties of isotropic variograms that will be referred to numerous times in the application sections to follow.

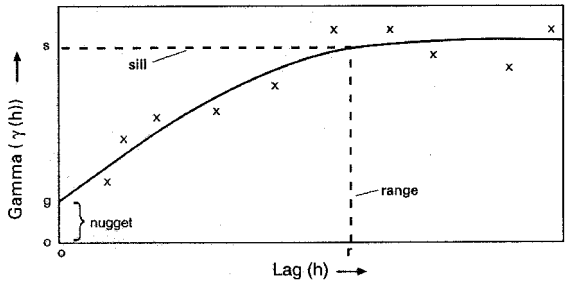
e. A plot of the sample variogram versus *h* often has a considerable degree of scatter (Figure 2-2), which is especially evident if the sample size *n* is small. However, the points can usually be fitted by a smooth curve that represents a theoretical variogram selected from a suite of possible choices. Usually, the theoretical variogram is monotonically increasing, signifying that the farther two observations are apart, the more their residuals tend to differ, on average, from one another. Several properties common to many theoretical variograms are shown in Figure 2-2. If the variogram either reaches or becomes asymptotic to a constant value as *h* increases, that value is called the **sill** (Figure 2-2). The distance (value of *h*) after which the variogram remains at or close to the sill is called the **range**. Measurements whose locations are farther apart than the range all have the same degree of association. Often, a variogram will have a discontinuity at the origin, signifying that even measurements very close together are not identical. Such variation in the measurements at small scales is called the **nugget effect**. The size of the discontinuity is called the nugget. Although the nugget effect is sometimes confused with measurement error, there is a subtle difference between these two concepts that will be explained in section 2-4. A simple monotonic function is usually selected to approximate the variogram. Four such functions that are often used in practice are:

the **exponential variogram** (parameters: sill, $s > 0$; nugget, $0 < g < s$; range, $r > 0$)

$$\gamma(h) = \begin{cases} g + (s - g)\left[1 - \exp\left(-3\frac{h}{r}\right)\right], & h > 0 \\ 0, & h = 0 \end{cases} \qquad (2\text{-}23)$$

the **spherical variogram** (parameters: sill, $s > 0$; nugget, $0 < g < s$; range, $r > 0$)

$$\gamma(h) = \begin{cases} s, & h > r \\ g + (s - g)\left[1.5\frac{h}{r} - 0.5\left(\frac{h}{r}\right)^3\right], & 0 < h \le r \\ 0, & h = 0 \end{cases} \qquad (2\text{-}24)$$



NOTE: The x's denote hypothetical sample variogram points computed from observed data. The smooth curve represents a theoretical variogram fitted to the sample variogram points.

Figure 2-2. Diagram showing variogram and features

the **Gaussian variogram** (parameters: sill, $s > 0$; nugget, $0 < g < s$; range, $r > 0$)

$$\gamma(h) = \begin{cases} g + (s-g)\left[1 - \exp\left(-3\left(\frac{h}{r}\right)^2\right)\right], & h > 0 \\ 0, & h = 0 \end{cases} \quad (2\text{-}25)$$

and, the **linear variogram** (parameters: nugget, $g > 0$; slope, $b > 0$)

$$\gamma(h) = \begin{cases} g + bh, & h > 0 \\ 0, & h = 0 \end{cases} \quad (2\text{-}26)$$

f. Although there are many other models that are used for variograms (Journel and Huijbregts 1978), these four are the most commonly used and are shown in Figure 2-3. The exponential, spherical, and Gaussian models are similar in that they all have a sill and a range. However, they have different shapes near zero lag (h=0) that, as will be discussed in Chapter 4, result in significant differences in the prediction results using the three models. The linear model is quite different from the other three, in that it does not reach a sill, but increases linearly without. This fact will have important implications on the prediction results using a linear variogram. Because the squared differences between residuals tend to increase without bound as the lag increases, a regionalized random variable with a linear variogram will have ever-increasing variability about its mean as the size of the sampling region is increased. In applications involving the linear variogram, the variogram is usually truncated at a sill corresponding to the value of the variogram at maximum lag h_{max}.

g. Before closing this section, it will be useful to highlight some similarities and contrasts between the covariance function and the variogram. Although the variogram is commonly used in a geostatistical analysis, it is sometimes easier to gain an intuitive understanding of the methodology using the covariance function, or equivalently, the spatial variance and the correlation function. When $Z(\underline{x})$ has a stationary, isotropic covariance function (Equation 2-9), there is a one-to-one correspondence between the variogram and the covariance function, namely

$$\gamma(h) = C(0) - C(h) \quad (2\text{-}27)$$

As long as $C(h)$ approaches zero as h increases (a minor technicality that can always be assumed in practice), then, as indicated by Equation 2-27, the variogram reaches a sill and the sill equals $C(0)$. Therefore, when dealing with a covariance-stationary regionalized random variable, the variogram and the spatial covariance function contain the same information as one another. By factoring out $C(0)=s$ from Equation 2-27 and using Equation 2-14, the relationship between the spatial correlation function and the variogram can be obtained

$$\rho(h) = 1 - \frac{\gamma(h)}{s} \quad (2\text{-}28)$$

From Equation 2-28, it is evident that high values of $\gamma(h)$ (i.e., close to s) signify low values of $\rho(h)$. In fact, $\rho(h) = 0$ whenever $\gamma(h) = s$, indicating that observations whose locations are farther apart than the range are uncorrelated. As h gets small, a nugget in $\gamma(h)$ is reflected in a correlation that is less than 1

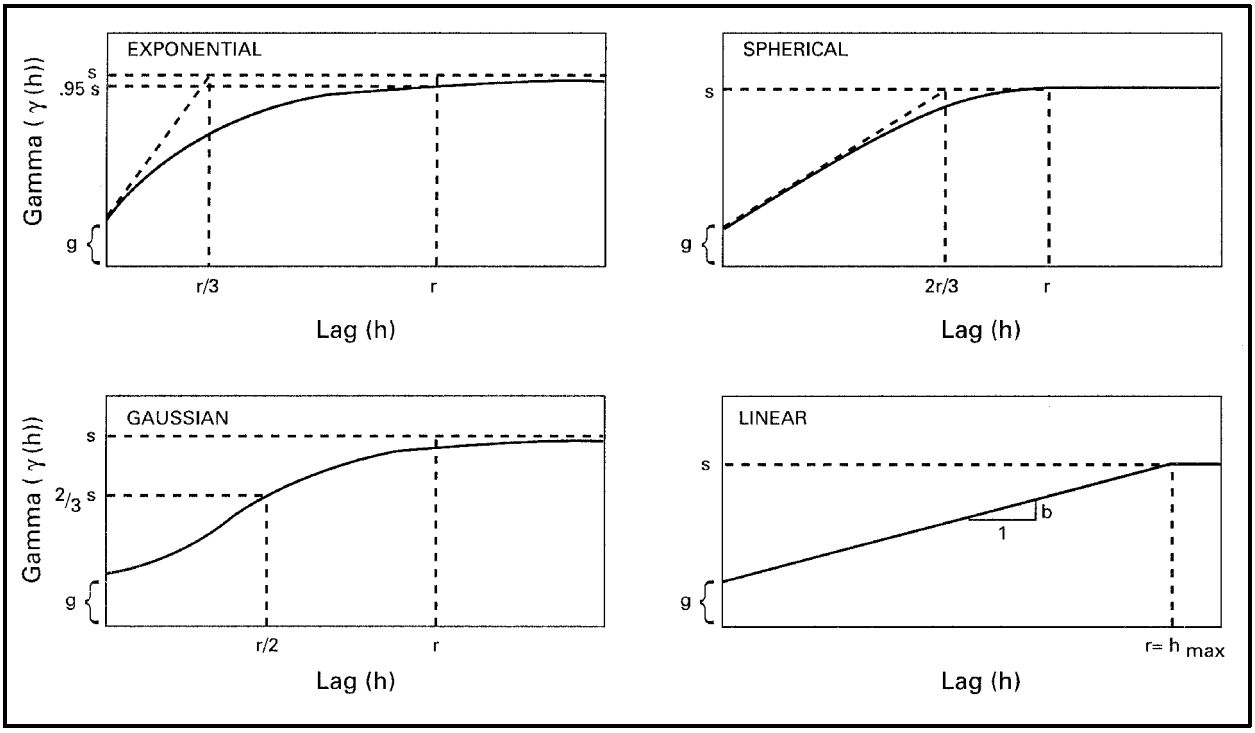


Figure 2-3. Theoretical variograms showing A, exponential; B, spherical; C, Gaussian; and D, linear models

$$\rho(h) \rightarrow 1 - \frac{g}{s} \text{ as } h \rightarrow 0 \qquad (2\text{-}29)$$

Therefore, the larger g is in relation to s, the less correlated nearby observations are. The case when $g=s$, called a **pure nugget variogram**, results in $\rho(h)=0$ for all $h>0$. In that case, neighboring observations are uncorrelated no matter how closely they are spaced.

h. Occasionally, $\gamma(h)$ may not reach a finite sill, as in the linear variogram Equation 2-26. In that case, it is not possible to define a correlation function as in Equation 2-28. The corresponding regionalized random variable is said to be intrinsically stationary (Journel and Huijbregts 1978), which is more general than covariance stationarity. The theory behind intrinsically stationary variograms will not be presented in this ETL. As long as a "pseudo-range" h_{max} is defined, all of the computations described below can be generalized.

2-4. Kriging

a. General.

(1) Given a regionalized random variable $Z(\underline{x})$ with a known theoretical variogram, the question is: how can the value of $Z(\underline{x})$ be predicted at an arbitrary location, based on measurements taken at other locations? Suppose that Z is measured at n specified locations: $Z(\underline{x}_1)$, ..., $Z(\underline{x}_n)$. For example, Z could correspond to hydraulic conductivity and the locations might correspond to n preexisting wells in an aquifer. Let a new location be given by $\underline{x}_0=(u_0,v_0)$ and denote the ith measurement location by $\underline{x}_i=(u_i,v_i)$. Suppose that, based on prior knowledge of the geology, there are no prevailing trends in hydraulic conductivity, so the mean of $Z(\underline{x})$ is assumed to be constant over the entire region:

$$\mu(\underline{x}) = \mu \text{ (constant)} \qquad (2\text{-}30)$$

(2) Suppose the investigator wants to predict the value of $Z(\underline{x}_0)$ by using a **linear predictor**, $\hat{Z}(\underline{x}_0)$, which is defined as a weighted linear combination of the measured data

$$\hat{Z}(\underline{x}_0) = \sum_{i=1}^{n} w_i Z(\underline{x}_i) \tag{2-31}$$

where w_i is the weight assigned to $Z(\underline{x}_i)$. To determine specific values for the weights, some criteria need to be specified for $\hat{Z}(\underline{x}_0)$ to be a good predictor of $Z(\underline{x}_0)$. The first criterion is that $\hat{Z}(\underline{x}_0)$ be an **unbiased predictor** of $Z(\underline{x}_0)$, which is expressed as

$$E[\hat{Z}(\underline{x}_0) - Z(\underline{x}_0)] = 0 \tag{2-32}$$

(3) An unbiased predictor will neither consistently overpredict nor underpredict $Z(\underline{x}_0)$ because the statistical expectation of the prediction errors is zero. The second criterion for a good predictor is that it have small **prediction variance**, defined by

$$\begin{aligned} &\mathrm{Var}\left[\hat{Z}(\underline{x}_0) - Z(\underline{x}_0)\right] \\ &= E\left[\left(\hat{Z}(\underline{x}_0) - Z(\underline{x}_0)\right)^2\right] \end{aligned} \tag{2-33}$$

(4) The smaller the prediction variance, the closer $\hat{Z}(\underline{x}_0)$ will be (on average) to the true value $Z(\underline{x}_0)$. The geostatistical method of kriging deals with computing the **best linear unbiased predictor** of $Z(\underline{x}_0)$, which is the linear unbiased predictor (Equations 2-31 and 2-32) with the smallest possible prediction variance (Equation 2-33).

(5) The form of the best linear unbiased predictor will depend on the mean of $Z(\underline{x})$. For example, if $Z(\underline{x})$ has a constant mean (Equation 2-30) and a pure nugget variogram [$\gamma(h)=s$ for all $h>0$], the best linear unbiased predictor of $Z(\underline{x}_0)$ will simply be the average of the measured data

$$\hat{Z}(\underline{x}_0) = \frac{1}{n} \sum_{i=1}^{n} Z(\underline{x}_i) \tag{2-34}$$

Because the variogram is the same for all $h>0$ and there is no trend in the data, there is no reason to favor any of the measurements over any of the other measurements. Therefore, the weights are all the same. Ordinary kriging, which is discussed in section 2-4*b*, deals with the constant-mean model (assumption in Equation 2-30) in which the variogram is not a pure nugget variogram. The weights of the best linear unbiased predictor will reflect the information in the variogram and will result in an improved predictor over the sample mean. In section 2-4*c*, universal kriging, which is the extension of ordinary kriging to the case of a nonconstant mean, is discussed. Universal kriging is a very powerful tool that can be used to combine regression models and spatial prediction into one unifying theory. Other, more specialized types of kriging that will be discussed in this section are indicator kriging (section 2-6*c*), block kriging (section 2-4*d*), and co-kriging (section 2-5).

(6) Before giving the kriging equations, one final note is in order. There is a prediction technique in geostatistics known as **simple kriging**, which deals with best linear unbiased prediction in the case when the mean of $Z(\underline{x})$ is fixed and known. Simple kriging is not discussed in this ETL, because, in most applications, the mean is not known and has to be estimated.

b. Ordinary kriging.

(1) General.

(a) Let $Z(\underline{x})$ be a regionalized random variable with constant mean (Equation 2-30) and isotropic variogram (Equation 2-20). Also, assume that the variogram reaches a sill so that the variance of $Z(\underline{x})$ is $C(0)=s$, and the correlation function is given by Equation 2-28. Although the prediction equations can be expressed in terms of the variogram, they will be defined here in terms of the sill (variance) and the correlation function.

(b) Consider linear unbiased predictors of the form of Equation 2-31 with the condition in Equation 2-32 holding. The unbiased condition is

equivalent to $\mu \sum_{i=1}^{n} w_i = \mu$ for any μ, which holds if and only if $\sum_{i=1}^{n} w_i = 1$. Therefore, all linear unbiased estimators need to have weights that sum to one. There are many sets of weights that satisfy this condition, including the set with all the weights equal to $1/n$, as in the sample mean, Equation 2-34. However, the unique set of weights that minimize the prediction variance (Equation 2-33) can be shown to satisfy the following set of $n+1$ **ordinary kriging equations** (Chapter 12, Isaaks and Srivastava (1989)):

$$\sum_{j=1}^{n} w_j \rho_{ij} + \frac{\lambda}{s} = \rho_{i0}, \quad i=1,2...,n, \tag{2-35a}$$

$$\sum_{j=1}^{n} w_j = 1 \tag{2-35b}$$

where $\rho_{ij} = \rho(h_{ij})$ is the correlation between observations i and j, h_{ij} is the distance between locations i and j, and λ is a coefficient resulting from the constrained optimization. Furthermore, the resulting **ordinary kriging variance** is

$$\begin{aligned} \sigma_k^2(\underline{x})_0 &= E\left[\left(\hat{Z}(\underline{x}_0) - Z(\underline{x}_0)\right)^2\right] \\ &= s\left(1 - \sum_{j=1}^{n} w_j \rho_{j0}\right) - \lambda \end{aligned} \tag{2-36}$$

(c) The system of Equations 2-35a and 2-35b can easily be solved for the w_i's and λ, after which the kriging variance can be obtained from Equation 2-36. Note that the ordinary kriging variance changes depending on the prediction location $\underline{x}_0$, even though the variance of $Z(\underline{x}_0)$ itself (Equation 2-6) is constant for all $\underline{x}_0$.

(2) Example 1.

(a) Let the mean of $Z(\underline{x})$ satisfy Equation 2-30, and suppose that the residual $Z^*(\underline{x})$ (Equation 2-16) has an isotropic exponential variogram (Equation 2-23). Consider predicting $Z(\underline{x}_0)$ based on $n=2$ measurements $Z(\underline{x}_1)$ and $Z(\underline{x}_2)$, where the three locations ($\underline{x}_0$, $\underline{x}_1$, and $\underline{x}_2$) are distinct. Using Equations 2-23 and 2-28, note that the correlation function is

$$\rho(h) = \begin{cases} \left(1 - \frac{g}{s}\right) \exp\left(-3\frac{h}{r}\right), & h>0 \\ 1, & h=0 \end{cases} \tag{2-37}$$

For illustrative purposes, suppose that

$$\frac{g}{s} = p, \quad 0 \le p \le 1 \tag{2-38}$$

where p is a fixed proportion. The quantity p is sometimes referred to as a **relative nugget**.

(b) The ordinary kriging Equations 2-35a and 2-35b are given by

$$w_1 + w_2 \rho_{12} + \frac{\lambda}{s} = \rho_{10} \tag{2-39a}$$

$$w_1 \rho_{12} + w_2 + \frac{\lambda}{s} = \rho_{20} \tag{2-39b}$$

$$w_1 + w_2 = 1 \tag{2-39c}$$

These three equations have three unknowns: w_1, w_2, and λ; the solution is

$$w_1 = \frac{1}{2} + \frac{1}{2}\frac{\rho_{10} - \rho_{20}}{1 - \rho_{12}} \tag{2-40a}$$

$$w_2 = \frac{1}{2} - \frac{1}{2}\frac{\rho_{10} - \rho_{20}}{1 - \rho_{12}} \tag{2-40b}$$

and

$$\lambda = \frac{s}{2}(\rho_{10} + \rho_{20} - \rho_{12} - 1) \tag{2-41}$$

The resulting kriging variance is

$$\sigma_K^2(\underline{x})_0 = s\left[\frac{3}{2} - w_1\rho_{10} - w_2\rho_{20} - \frac{1}{2}\left(\rho_{10} + \rho_{20} - \rho_{12}\right)\right] \quad (2\text{-}42)$$

Although there are only three sample locations in this example (two actual and one potential), it indicates several properties of best linear unbiased prediction that hold in general. For example,

(c) **Effect of sill**. The kriging weights depend on *s* only through the relative nugget *p*. However, the kriging variance is directly proportional to *s*. The sill is called a scaling parameter because scaling each measurement by a constant *c* has the effect of scaling *s* by c^2. When the relative nugget is allowed to vary so that *s* and *g* can change independently, the effect of *s* is somewhat more complicated.

(d) **Effect of nugget**. Increasing *p* has the effect of drawing each of the weights closer to 1/2. In fact, as *p* approaches 1, both weights will equal 1/2. The larger *g* is in relation to *s*, the more small-scale variability there is in the data and the less important the correlation between neighboring locations becomes. The increased small-scale variability also causes an increase in the kriging variance.

(e) **Effect of correlations**. If $Z(\underline{x}_0)$ is more highly correlated with $Z(\underline{x}_1)$ than with $Z(\underline{x}_2)$, then w_1 will be larger than w_2, indicating that the measurement at the first location has more predictive information than the measurement at the second location. Also, correlation in the data always decreases the kriging variance compared to the variance with uncorrelated data.

(f) **Effect of data clumping**. If $Z(\underline{x}_1)$ and $Z(\underline{x}_2)$ are highly correlated, as indicated by ρ_{12} being close to 1, then the two measurements contain much of the same information. Two situations can occur: $\rho_{10} = \rho_{20}$, in which case the weights are both equal, or $\rho_{10} > \rho_{20}$ [$\rho_{10} < \rho_{20}$], in which case w_1 will be much larger [smaller] than w_2. In either case, the kriging variance will increase to reflect the redundant information in the two measurements. Automatic adjustment of the kriging weights and kriging variance to account for data clumping is an important property of the kriging predictor.

(3) Example 2 (Nugget effect versus measurement error).

(a) In example 1, all three locations $\underline{x}_0$, $\underline{x}_1$, and $\underline{x}_2$, were assumed to be distinct. When a prediction location happens to coincide with a measurement location, there is an important distinction that needs to be made between a true nugget effect and a measurement error. Suppose that in example 1, $\underline{x}_0$ and $\underline{x}_1$ are the same. If there is only small-scale variability, but no measurement error, then repeated measurements at the same location should be identical, that is, $\rho_{10} = 1$. In this case, the kriging equations result in $w_1 = 1$, $w_2 = 0$, and $\lambda = 0$ and in a kriging variance of zero. That is, $Z(\underline{x}_1)$ is a perfect predictor of $Z(\underline{x}_0)$. This property, called **exact interpolation**, is a property of kriging when the data are assumed to contain no measurement errors. However, suppose instead that the nugget is interpreted as measurement error rather than small-scale variability. In that case, repeated measurements at the same location would not be perfectly correlated, but rather, $\rho_{10} = 1\text{-}g/s$.

(b) Substituting this correlation into the kriging equations and solving the equations results in a predictor that does not exactly interpolate the data, but instead smooths the measured data to account for the measurement error. In this ETL, prediction locations are assumed not to coincide with measurement locations, in which case no distinction needs to be made between nugget and measurement error.

c. Universal kriging.

(1) Universal kriging is an extension of ordinary kriging, that, due to the fact that environmental data often contain drift, can be important in HTRW site investigations. Universal kriging addresses the case of a nonconstant mean $\mu(\underline{x})$.

Generally, the mean is assumed to have a functional dependence on spatial location of the form

$$\mu(u, v) = \sum_{j=1}^{p} \beta_j f_j(u, v) \qquad (2\text{-}43)$$

where the $f_j(u,v)$'s are known deterministic functions of $\underline{x}=(u,v)$ (that is, these functions serve as independent variables) and the β_j's are regression coefficients to be estimated from the data. For example, suppose $Z(\underline{x})$ is hydraulic head in an aquifer. If the flow is in a steady state, it might be reasonable to assume, in a given case, that the mean of $Z(\underline{x})$ has a unidirectional groundwater gradient that is given by

$$\mu(u, v) = \beta_1 + \beta_2 u \qquad (2\text{-}44)$$

In this example, there are two independent variables:

$$f_1(u, v) = 1$$
$$f_2(u, v) = u \qquad (2\text{-}45)$$

and two regression coefficients (β_1 and β_2). The mean can include other independent variables besides simple algebraic functions of u and v. For example, if the aquifer is not of uniform thickness, an independent variable that involves the aquifer thickness at location (u,v) could be included.

(2) The form assumed for the mean in Equation 2-43 is also generally used in standard linear regression analysis. In regression, ordinary least-squares is generally used to solve for the coefficients; when this is done, it is assumed that the residuals are independent and identically distributed. Universal kriging is an extension of ordinary least-squares regression that allows for spatially correlated residuals. Assuming that $Z(\underline{x})$ is a regionalized random variable with a mean as in Equation 2-43 and residual correlation function as in Equation 2-28, the best linear unbiased predictor (Equation 2-10) can be obtained from the following $n+p$ equations, called the **universal kriging equations** (Journel and Huijbregts 1978):

$$\sum_{j=1}^{n} w_j \rho_{ij} + \frac{1}{s} \sum_{k=1}^{p} \lambda_k f_k(\underline{x}_i) = \rho_{i0}, \quad i=1, 2, ..., n \qquad (2\text{-}46a)$$

$$\sum_{j=1}^{n} w_j f_k(\underline{x}_j) = f_k(\underline{x}_0), \quad k =, 1, 2, ..., p \qquad (2\text{-}46b)$$

where, in contrast to the ordinary kriging equations (2-35a and b), there are now p coefficients $\lambda_1, ..., \lambda_p$ resulting from the unbiased condition on the predictor. The first term in the mean (Equation 2-43) will usually be a constant, or intercept, for which $f_1(\underline{x}) = 1$. Therefore, the universal kriging model includes ordinary kriging as a special case. The **universal kriging variance** is given by

$$\sigma_k^2(\underline{x}_0) = s\left(1 - \sum_{i=1}^{n} w_i \rho_{i0}\right) - \sum_{k=1}^{p} \lambda_k f_k(\underline{x}_0) \qquad (2\text{-}47)$$

These equations can be easily solved to obtain universal kriging predictors and kriging variances for any desired location. The estimated trend surface does not actually need to be computed to obtain the universal kriging predictor. If a particular application needs an estimate of the trend surface, then generalized least-squares regression can be used to estimate the coefficients (β_j's) in the regression equation.

d. Block kriging.

(1) Up to this point, the problem of predicting the value of a regionalized random variable at a given location in the region over which the variable is defined has been considered. Implicit in this analysis is the assumption that the support of the variable being predicted is defined in exactly the same way as the variables that make up the measurements. However, there may be applications

where it is necessary to estimate the average value of Z over an estimation block of much larger area than is represented by an individual sample. For example, an estimate of the average concentration of a contaminant over an entire aquifer based on point measurements at various locations might be needed. In other applications, an estimate of the average concentration of soil contaminant in daily excavation volumes that are much larger than the volume of an individual sample may be needed. Let Z_B be the average value of $Z(\underline{x})$ over a particular block B,

$$Z_B = \frac{1}{m} \sum_{i=1}^{m} Z(\underline{x}_{0i}) \qquad (2\text{-}48)$$

where $\underline{x}_{0i}$, i=1,...,m, denotes m prediction locations in block B. The object is to predict this average rather than the regionalized variable at a single location. In many applications, the locations $\underline{x}_{0i}$ might correspond to nodes of a regular grid or finite- element nodes in a groundwater model. Results of the block kriging are dependent on m and on the placement of the prediction locations. Selecting a large number of locations in block B, where each location has approximately the same representative area, is the best approach (Chapter 13, Isaaks and Srivastava (1989).

(2) The objective of block kriging is to obtain the best linear unbiased predictor of Z_B and an estimate of the block kriging variance based on the measurements. The model for $Z(\underline{x})$ can be the constant-mean model (Equation 2-30) assumed for ordinary kriging or the more general linear regression model (Equation 2-43) assumed for universal kriging. In either case, the predicted value of Z_B coincides with the average of the predicted values of the individual measurements in the block; that is

$$\hat{Z}_B = \frac{1}{m} \sum_{i=1}^{m} \hat{Z}(\underline{x}_{0i}) \qquad (2\text{-}49)$$

In this equation, the individual predicted values are obtained from either the ordinary or universal kriging equations. However, computation of the block kriging variance is not as simple, because the individual kriging estimates are not independent of one another. There are simple modifications to the kriging equations discussed in sections 2-4*b* and 2-4*c* that can be used to directly compute the kriging estimate of Z_B, along with its kriging variance (Chapter 13, Isaaks and Srivastava (1989)). The equations are not presented in this ETL. The computer packages described in the next section can be used to compute block kriging estimates. In general, kriged values of block averages are less variable than kriged values at single locations. Consequently, the blocked kriging variance tends to be smaller than the kriging variance at a single location.

2-5. Co-kriging

a. Kriging as discussed so far provides a way of predicting values of a regionalized variable $Z(\underline{x})$ at a location $\underline{x}_0$ based on measurements of the same variable at locations $\underline{x}_1$, $\underline{x}_2$, ..., $\underline{x}_n$. In some situations, however, there will be available measurements not only of $Z(\underline{x})$, but also of one or more other variables that can be used to improve predictions of $Z(\underline{x}_0)$. The variable $Z(\underline{x})$ will be called the primary variable, because it is the one to be predicted, and the other variables will be called secondary variables. **Co-kriging** is the technique that allows the use of the information contained in secondary variables in the prediction of a primary variable. As an example, suppose that $Z(\underline{x})$ is a regionalized variable representing the hexavalent chromium concentration, a relatively difficult determination, and that hexavalent chromium concentration needs to be predicted at a location $\underline{x}_0$ based on measurements of hexavalent chromium at other locations, but there are also measurements of a second relatively easily determined contaminant, for example lead, that tend to be correlated with hexavalent chromium concentration and these data are to be used as well. Denote the second variable lead by a regionalized variable $W(\underline{x})$, and assume that measurements have been made on W at m locations $\underline{x}'_1$ $\underline{x}'_2$, ..., $\underline{x}'_m$. The co-kriging predictor of $Z(\underline{x})_0$ is then

$$\hat{Z}_C(\underline{x}_0) = \sum_{i=1}^{n} w_i Z(\underline{x}_i) + \sum_{j=1}^{m} w'_j W(\underline{x}'_j) \quad (2\text{-}50)$$

This is a straightforward extension of the kriging predictor in Equation 2-31. Analogous to kriging, co-kriging produces the weights w_i and w'_j so that the resulting predictor is the best linear unbiased predictor. Also, as with kriging, co-kriging requires modeling of the variogram for Z, but co-kriging presents the investigator with the additional necessity of modeling the variogram of W and the **cross variogram** for Z and W. The optimal weights are then expressed in terms of all these variogram properties. More than one secondary variable may be included in the co-kriging predictor, and theory has been developed for co-kriging in the presence of drift (universal co-kriging) and block co-kriging. Details are not included in this ETL, but the interested reader may refer to Isaaks and Srivastava (1989) and Deutsch and Journel (1992) for more discussion and citation of other references.

b. One situation in which co-kriging might be useful is when the primary variable is undersampled, so any additional information, such as that given by secondary variables, would be helpful. However, although co-kriging can be a useful tool, joint modeling of several variables tends to be demanding in terms of data and computational requirements. Thus, undersampling of the primary variable may present problems for co-kriging as well as for one-variable kriging. Also, unless the primary variable of interest is highly correlated with the secondary variable(s), the weights assigned to the secondary variable(s) are often small, with the result that the effort needed to include the additional variable(s) may not be worthwhile. For these reasons, co-kriging tends not to be used extensively in practice.

c. Although co-kriging is similar to universal kriging, in that both techniques use extra variables to help predict $Z(\underline{x})$, there is an important distinction between the two techniques. In universal kriging, the independent variables in Equation 2-43 need to be known with certainty at the prediction location $\underline{x}_0$. For example, aquifer thickness might be an independent variable in predicting aquifer head if it can easily be determined at any location. However, aquifer thickness may need to be considered a secondary variable in a co-kriging procedure if it is only known at a few selected points in the aquifer.

2-6. Using Kriging to Assess Risk

a. General.

(1) The kriging predictor of $Z(\underline{x}_0)$ has certain desirable properties with respect to how close it is to the actual value of $Z(\underline{x}_0)$, it is unbiased and has smallest variance among all linear predictors. On the average, or in an expected sense, the predicted value will be near the actual value. When possible, however, the investigator would like to go further in specifying the relationship between the predicted and observed values. Ideally, the investigator would like to make probability statements. For example, if $Z(\underline{x}_0)$ is concentration of a contaminant, the investigator might like to be 95 percent certain that the true concentration is within 0.05 ug/ℓ of the predicted concentration. In other situations, the probability that the actual concentration exceeds a given target value might need to be estimated. Knowledge of the entire distribution function of $Z(\underline{x})$, as opposed to knowledge of only its mean and variogram, can be used for risk-qualified inferences in situations when extremes might be of more interest than averages.

(2) Introduction of the concept of a **conditional probability distribution** function of the regionalized variable $Z(\underline{x})$ is appropriate at this point. This concept will also be used in Chapter 7 when conditional simulation is discussed. The conditional probability distribution function has a definition much like that of the probability distribution function in section 2-2, except the probability that $Z(\underline{x}) \leq c$ is computed "conditional on," or "given," information at other spatial locations. The interest in geostatistics is to make predictions

at a location $\underline{x}_0$ using information at measurement locations $\underline{x}_1, \underline{x}_2, ..., \underline{x}_n$, so, in terms of conditional distributions, interest focuses on $P[Z(\underline{x}_0) \le c \mid Z(\underline{x}_1), Z(\underline{x}_2), ..., Z(\underline{x}_n)]$. The vertical bar denotes the conditioning and is read "given." This conditional probability distribution needs to be determined to make probability statements about the regionalized variable at location $\underline{x}_0$. Also, **conditional mean** and **conditional variance** can be defined in the present context in the same way that mean and variance for distribution functions were defined in section 2-2.

(3) Section 2-6*b* contains methods for using kriging output to obtain prediction intervals or quantiles when the regionalized random variable is either normally distributed or can be transformed to a near-normal distribution. Section 2-6*c* discusses indicator kriging, which is a nonparametric method for obtaining quantiles when data cannot be transformed adequately to a normal distribution.

b. Normal distributions and transformations.

(1) For prediction at a location $\underline{x}_0$, a kriging analysis produces the predictor $\hat{Z}(\underline{x}_0)$ and the associated kriging variance $\sigma_k^2(\underline{x}_0)$. If more informative probability assessments are to be made, the ideal situation is when $Z(\underline{x})$ can be assumed to be a Gaussian, or normal, process, which means that $[Z(\underline{x}_1),..., Z(\underline{x}_n)]$ has a joint normal probability distribution for any set of n locations and any value of n. In this case, the conditional probability distribution of $Z(\underline{x}_0)$ given the n observations is a normal distribution with conditional mean equal to the kriging predictor $\hat{Z}(\underline{x}_0)$ and conditional variance equal to the kriging variance $\sigma_k^2(\underline{x}_0)$. This normal distribution can be used to obtain a **prediction interval** for $Z(\underline{x}_0)$ (conditional on the measured data). For example, from a table of the normal distribution, a value of 1.96 corresponding to a 0.95 (two-sided) probability can be obtained. Then the assertion that there is a 95-percent chance that $Z(\underline{x}_0)$ will be in the 95-percent prediction interval $[\hat{Z}(\underline{x}_0) - 1.96\ \sigma_K(\underline{x}_0),\ \hat{Z}(\underline{x}_0) + 1.96\ \sigma_K(\underline{x}_0)]$ can be made. Knowing this interval is much more useful than simply knowing the kriging predictor and variance.

(2) To illustrate quantile estimation, suppose that contaminant concentrations are being studied and the concentration that has only a 1-percent chance of being exceeded at location $\underline{x}_0$ needs to be determined. The appropriate (one-sided) value from a normal table is 2.33, so the desired estimate is $\hat{Z}(\underline{x}_0) + 2.33\sigma_K(\underline{x}_0)$.

(3) Even if $Z(\underline{x})$ is not Gaussian, it is often possible to find a **transformation**, $Y(\underline{x})=T(Z(\underline{x}))$, such that $Y(x)$ is approximately Gaussian. When a transformation is made, the kriging analysis is performed using the transformed data $Y(\underline{x})$, and the inverse transformation may be applied to obtain prediction intervals for the original data. For example, the most common transformation is the (natural) **logarithmic transformation**, in which $Y(\underline{x})=1n[Z(\underline{x})]$. A 95-percent prediction interval for $Z(\underline{x})$ is then $\{\exp[\hat{Y}(\underline{x}_0) - 1.96\ \sigma_k(\underline{x}_0)], \exp[\hat{Y}(\underline{x}_0) + 1.96\ \sigma_k(\underline{x}_0)]\}$. As long as the transformation is a one-to-one function such as a logarithmic transform, prediction intervals for the original data can be obtained by simply back-transforming prediction intervals for the transformed data.

(4) Although it is a simple matter to obtain prediction intervals and probabilities using simple back-transformation, it is more difficult to obtain a predictor of the untransformed data that is both unbiased and optimal in some sense. For example, in the case of a logarithmic transformation, a kriging analysis using the transformed data yields a predictor $\hat{Y}(\underline{x}_0)$, which is the best linear unbiased predictor of $Y(\underline{x}_0)$. However, the back-transformed value $\hat{Z}(\underline{x}_0) = \exp[\hat{Y}(\underline{x}_0)]$ does not possess these same optimality properties as a predictor of $Y\underline{x}_0$. The methodology known as log-normal kriging, and more generally trans- normal kriging, has been developed to obtain predictors in this setting (Journel and Huijbregts 1978), but because of the complexity involved in these procedures, they are not usually used by practitioners. If a predicted value corresponding to $Z(\underline{x}_0)$ needs to be obtained for purposes of contour plotting, the kriging predictions $\hat{Y}(\underline{x}_0)$ may be back-transformed and plotted, as long as the investigator realizes that such values do not have the usual kriging optimality properties.

c. Indicator kriging.

(1) There may be situations when a transformation that makes $Z(\underline{x})$ approximately normal cannot be easily determined. In such situations, **indicator kriging** can be used to make inferences about the probability distribution of $Z(\underline{x})$. Because no distributional assumptions are made, this technique is known as a **nonparametric** statistical procedure. An example of indicator kriging is included in Chapter 5, and a paper by Journel (1988) is a good reference for additional information about indicator kriging.

(2) To perform indicator kriging, a special transformation, known as an indicator transformation, is applied to $Z(\underline{x})$:

$$I(\underline{x},c) = \begin{cases} 1, & Z(\underline{x}) \le c \\ 0, & Z(\underline{x}) > c \end{cases} \qquad (2\text{-}51)$$

If, as in the usual kriging scenario, the data set at hand consists of measurements of the regionalized variable $Z(\underline{x})$ at n locations, c needs to be fixed first, and then the indicator transformation is applied by replacing values that are less than or equal to c with 1 and values that are greater than c with 0. The variogram and kriging analysis is then performed using these 0's and 1's rather than the raw data.

(3) Kriging predictors using the indicator data will be equal to their observed values of 0 or 1 at the measurement locations $\underline{x}_i$, i=1,...,n. However, at locations different from the measurement locations, predictions may be between 0 and 1. In interpreting these values, the power of indicator kriging becomes apparent. A predicted value at $\underline{x}_0$ is an estimate of the conditional probability distribution $P[Z(\underline{x}_0) \le c \mid Z(\underline{x}_1)\ Z(\underline{x}_2), ..., Z(\underline{x}_n)]$. This analysis may be performed for a range of values of c, and by doing this the entire distribution function can be estimated. This estimate of the distribution function can be used in the same manner discussed above to obtain prediction intervals or estimates of quantiles. For example, to estimate the value that has a 1-percent chance of being exceeded at location $\underline{x}_0$, the value of c for which the kriged indicator prediction is 0.99 at that location is determined.

(4) One advantage of indicator kriging is that the indicator variogram is robust with respect to extreme outliers in the data because no matter how large (or small) $Z(\underline{x})$ is, the indicator variable is either 0 or 1. Indicator variables may also be used in the context of block kriging. For example, a spatial average of $I(\underline{x},c)$ over a block B equals the fraction of block B for which $Z(\underline{x})$ is less than c. Another advantage of indicator kriging is that it can be used when some data are censored.

(5) Despite the relative ease of implementation, there are several drawbacks to indicator kriging, and investigators may wish to use this technique only when other methods, such as normality transformations, produce unacceptable results. For example, the kriged values of $I(\underline{x},c)$ may be less than 0 or larger than 1. Also, the kriged prediction for $I(\underline{x},c_1)$ may be larger than the kriged prediction for $I(\underline{x},c_2)$ even if $c_1<c_2$, which is not compatible with a valid probability distribution. There are several more advanced techniques for dealing with these problems (Chapter 18, Isaaks and Srivastava (1989); however, they are beyond the scope of this ETL.